Evaluating the global testing culture

Evaluating the global testing culture: a response to Ramirez and colleagues

The recently published article by Ramirez and colleagues (2018), *International Tests, National Assessments, and Educational Development (1970-2012,)* is unique in using panel data from more than four decades to evaluate the potential negative consequences of assessment. Central to their analysis is recent work on the global testing culture (Smith, 2016a); demonstrated in the overarching research question "does a global testing culture dominate the global educational regime to the detriment of educational outcomes favored by most reformers and generally regarded as progress?" (Ramirez et al., 2018, p. 349). While acknowledging the presence of a global testing culture, the authors conclude that their findings "do not support these dire predictions" (Ramirez et al., 2018, p. 344) of testing critics. Instead of negative outcomes, they find no relationship between the number of international and national assessments countries partake in and measures of secondary school enrollment and gender parity and a positive association between international assessments and outcomes at the tertiary level. In addition, greater participation in international tests are related to more progressive content in social science textbooks. As testing critic, I do not find these findings at all surprising. While Ramirez and colleagues expand our understanding of world culture more generally, they do not effectively evaluate the concerns from those worried about the global testing culture. As I suggest in this response, the test categories and outcome variables chosen in this analysis fail to capture key parts of such arguments. In the proceeding sections, I clarify the global testing culture and its associated critiques before suggesting fruitful areas for future research.

The global testing culture and types of testing

The global testing culture (Smith, 2016b) is "characterized by census-based standardized testing with links to high stakes outcomes" (p. 7). In this culture such tests are "accepted as foundational practice in education and shape how education is understood in society and used by its stakeholders" (p. 10). It has emerged, in part, as a result of high stakes associated with testing being increasingly transferred from students to teachers and schools.

Ramirez and colleagues are among a growing number (Harris & Harrington, 2006; Lingard et al., 2013; UNESCO, 2018) who rightly recognize that "different types of tests are associated with various educational outcomes" (p. 358). Unfortunately, their categorization limits the analysis and conclusions that can be drawn in relation to the global testing culture. Examining test purpose is one of many ways tests can be grouped or classified (Smith, 2014a; UNESCO, 2018). For example, high stakes tests for advancement are used to decide whether students continue their education or in which track they can proceed. Alternatively, testing for accountability recognizes that at least part of the purpose of some tests is to hold schools and educators accountable via student test scores (Smith, 2014a). While testing for advancement was acknowledged in the article as national selection exams, before ultimately being removed from analysis, the second purpose of testing was forgotten about altogether.

Part of this may be due to differing definitions of high stakes. The authors recognize high stakes tests as "the presence of nationwide exams, typically administered by the education ministry, which determine student access to upper levels of educational system and/or to sharply differentiated education tracks" (Ramirez et al., 2018, p. 349). This mirrors the traditional understanding of high stakes exams, originally present in China (Eckstein & Noah, 1993) and still used in some countries (for Europe see Eurydice, 2009; for Asia see Hill, 2010). Nevertheless, limiting high stakes to students contrasts the broader literature on the

global testing culture which recognizes high stakes as a motivational lever applicable to multiple actors in education. For instance, Chapman and Snyder (2000) identify stakes as high stakes if they have "real or perceived consequences for students, staff, or schools" (p. 458). Specific to teachers, Larsen (2005) suggests stakes are high when results are "tied to increases in salary, promotion, and maintenance of employment" (p. 296).

Understanding the transformation of stakes from students to schools and educators is essential when evaluating whether the global testing culture has negative consequences for society. Instead, the article does not consider high stakes options beyond national selection exams. Once those are removed from the analysis, international assessments and national assessments are left. Without distinguishing how national assessments are used for accountability purposes, the applied classification scheme conflates sample based low stakes national assessments with census based high stakes testing for accountability. By combining these tests into a single category it is impossible to effectively interpret the results related to national assessments.

Clarifying the critiques of testing for accountability

In examining whether the global testing culture has led "to the detriment of educational outcomes" (Ramirez et al., 2018, p. 349), it is important to clearly disentangle the critiques. Highlighted undesirable consequences associated with practices of the global testing culture include narrowing the curriculum, teaching to the test, focusing on 'bubble' students or those at the margin of passing, and shaping the testing pool by transferring students into non-tested student groups or selectively admitting or transferring students depending on their achievement level (UNESCO, 2017).

Ramirez et al. (2018) attempts to evaluate claims of narrowing the curriculum and shaping the testing pool. Content in social science textbooks are used by the authors to evaluate whether increased testing has narrowed the curriculum in countries. However, this does not align with the arguments laid out by testing critics. For them, narrowing the curriculum involves shifting instructional time and resources toward tested subjects and away from non-tested subjects (Retner et al., 2006; Supovitz, 2009). Results suggesting that social science curriculum is becoming more progressive do not evaluate narrowing the curriculum in this sense. Furthermore, as social science is a less commonly tested subject (Benavot & Koseleci, 2015), it is perhaps expected that more student centered instruction and progressive content is found in these textbooks.

Similarly, dependent variables used to examine whether testing shapes the student population do not directly address concerns over who is included in the testing pool. When schools shape the testing pool they may not necessarily be reducing enrollment. Instead, they may transfer students into special education (Anderson & Boyle, 2015) or disproportionately discipline low performers during testing periods (Figlio, 2006). As manipulating those included in the testing pool is based on past and expected student achievement, it is unclear why the authors suggest women would participate at a different rate without having more country specific cultural context. Finally, repetition rates are a potential variable of interest. Research, however, suggests that students are increasingly held back in the year prior to the test (Haney, 2000). Therefore, without specific information on the tested grade, this analysis, which combines all primary or secondary repeaters into a single level, does not validly test the critique.

Evaluating the global testing culture

Future research hoping to examine the consequences of the global testing culture should, at a minimum, include testing for accountability when categorizing types of tests and evaluate critiques against the corresponding outcomes. Capturing testing for accountability, however, is a resource intense task that requires country by country analysis of policies with changes tracked over time. National testing policies (NTPs) have been put forth as one potential solution (Smith, 2014a). NTPs further divide testing for accountability into evaluative policies – those which aggregate student test scores by school and make them publicly available – and punitive policies – those that use school aggregate test scores to reward or punish schools. Initial country categorization included participants of the 2009 PISA (Smith, 2017); UNESCO (2017) updated the list in 2017 to include over 100 countries.

Two promising areas for further research include using NTPs to examine narrowing the curriculum and shaping the testing pool. A suggested indicator for the prior would be instructional time; questioning whether testing for accountability countries spend more of their instructional time on tested subjects. Initial research suggests this may be true as students in punitive and evaluative countries spend an extra 44 and 13 minutes a week on mathematics, respectively, compared to those not in a testing for accountability country (Smith, 2014b). Some might suggest such countries may simply have a greater amount of total instructional time. However, data on compulsory instructional time from the OECD's Education at a Glance (OECD, 2011) for the same age group during the same year (although a slightly reduced country sample) suggests that non-testing for accountability countries had between 145 and 209 more minutes of instructional time each week.

Analysis of within country inequality can more easily get to the heart of shaping the testing pool claims. The global testing culture does not suggest children are omitted entirely from the school system. Instead, it would expect to find achievement based segregation in countries that practice testing for accountability. This appears to be the case in some countries, such as the United States (Vogell & Fresques, 2017) and Sweden (Osth et al., 2013), where lower performing students are increasingly concentrated together in low performing, poorly resourced schools.

Finally, it is also important to look beyond education. The global testing culture suggests all actors associated with education are effected "including both direct actors – students, teachers, parents, and government – and indirect actors – businesses, investors, and the larger community" (Smith, 2016b, p. 14). This opens an array of potential questions, such as how donors include test scores in funding decisions for development projects or how communities use test scores as an indicator of low school quality.

References

Anderson, J., & Boyle, C. (2015). Inclusive education in Australia: Rhetoric, reality and the road ahead. *Support for Learning*, *30*(1), 4–22.

Benavot, A. & Koseleci, N. (2015). Seeking quality in education: the growth of national learning assessments, 1990-2013. Background paper for Education for All Global Monitoring Report 2015. Paris: France.

Chapman, D., & Snyder, C. (2000). Can high stakes national testing improve instruction: re -examining conventional wisdom. *International Journal of Educational Development, 20*, 457-474.

Eckstein, M., & Noah, H. (1993). *Secondary School Examinations*. New Haven, CT: Yale University Press.

Eurydice. (2009). *National Testing of Pupils in Europe: Objectives, Organisation and use of Results*. Education, Audiovisual and Culture Executive Agency: European Commission.

Figlio, D. (2006). Testing, crime, and punishment. *Journal of Public Economics*, 90, 837-851.

Haney, W. (2000). The myth of the Texas miracle in education. Education Policy Analysis Archives, *8*(41).

Harris, D. & Herrington, C. (2006). Accountability, standards, and the growing achievement gap: lessons from the past half-century. *American Journal of Education,* 112(2), 209-238.

Hill, P. (2010). *Examination Systems: Asia-Pacific Secondary Education System Review Series.* Bangkok: UNESCO Bangkok.

Larsen, M.A. (2005). A critical analysis of teacher evaluation policy trends. *Australian Journal of Education, 49*(3), 292-305.

Lingard, B., Martino, W. & Rezai-Rashti, G. (2013). Testing regimes, accountabilities and education policy: commensurate global and national developments. *Journal of Education Policy*, *28*(5), 539-556.

OECD. (2011). Education at a Glance. Paris: OECD.

Osth, J., Andersson, E. & Malmberg, B. (2013). School choice and increasing performance difference: a counterfactual approach. *Urban Studies, 50*(2), 407-425.

Ramirez, F.O., Schofer, E. & Meyer, J. W. (2018). International tests, national assessments, and educational development (1970-2012). *Comparative Education Review, 62*(3), 344-364.

Rentner, D.S., Scott, C., Kober, N., Chudowsky, N., Chudowsky, V., Joftus, S. et al. (2006). *From the Capital to the Classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Center on Education Policy.

Smith, W.C. (2017) National testing policies and educator based testing for accountability: The role of selection in student achievement. *OECD Journal: Economic Studies, 2016/1*, 131-150.

Smith, W.C. (Ed.) (2016a). *The Global Testing Culture: Shaping Educational Policy, Perceptions, and Practice.* Oxford: Symposium Books.

Smith, W.C. (2016b). An introduction to the global testing culture. In Smith, W. (Ed.)**,** *The Global Testing Culture: Shaping Educational Policy, Perceptions, and Practice* (pp. 7-24). Oxford: Symposium Books.

Smith, W.C. (2014a). The global transformation toward testing for accountability. *Education Policy Analysis Archives, 22*(116).

Smith, W.C. (2014b). The global expansion of the testing culture: National testing policies and the reconstruction of education. Unpublished Ph.D. dissertation: The Pennsylvania State University.

Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change, 10,* 211-227.

UNESCO. (2018). *The Culture of Testing: Sociocultural Impacts on Learning in Asia and the Pacific.* Paris: UNESCO.

UNESCO. (2017). *Accountability in Education: Meeting our Commitments. 2017/8 Global Education Monitoring Report.* Paris: UNESCO.

Vogell, H. & Fresques, H. (2017). Hidden dropouts: how schools make low achievers disappear. *USA Today* (February 22, 2017). Accessed on August 26, 2018 at https://eu.usatoday.com/story/news/2017/02/21/hidden-dropouts-how-schools-game-system-dumping-underachievers-into-alternative-programs/97708866/